HHS Public Access

Author manuscript

J Proteomics. Author manuscript; available in PMC 2021 May 30.

Published in final edited form as:

J Proteomics. 2020 May 30; 220: 103777. doi:10.1016/j.jprot.2020.103777.

Experimentally-Driven Protein Structure Modeling

Nikolay V. Dokholyan[1]

[1]Department of Pharmacology, Penn State University College of Medicine, Hershey, Pennsylvania 17033, USA; Department of Biochemistry & Molecular Biology, Penn State College of Medicine, Hershey, Pennsylvania, 17033, USA; Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, USA; Department of Biomedical Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, USA

Abstract

Revolutions in natural and exact sciences started at the dawn of last century have led to the explosion of theoretical, experimental, and computational approaches to determine structures of molecules, complexes, as well as their rich conformational dynamics. Since different experimental methods produce information that is attributed to specific time and length scales, corresponding computational methods have to be tailored to these scales and experiments. These methods can be then combined and integrated in scales, hence producing a fuller picture of molecular structure and motion from the “puzzle pieces” offered by various experiments. Here, we describe a number of computational approaches to utilize experimental data to glance into structure of proteins and understand their dynamics. We will also discuss the limitations and the resolution of the constraints-based modeling approaches.

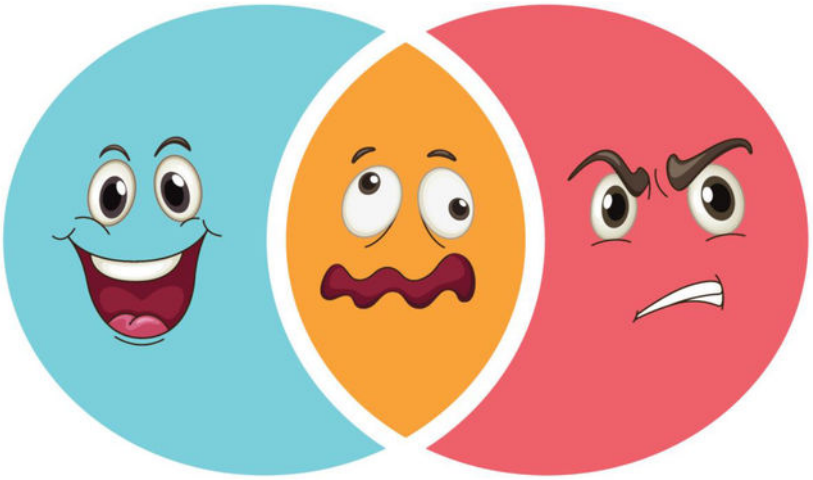
Graphical Abstract

Correspondence: dokh@psu.edu, Tel: (717) 531-5177.

Credit Author Statement

All of the work was performed by the sole author of the review article.

Conflict of interests: NONE



Keywords

molecular modeling; molecular dynamics simulations; discrete molecular dynamics; statistical mechanics; structural proteomics; mass spectrometry

Introduction

The past century has been transformational for many scientific disciplines. Some of the most prominent revolutions happened in biology: the field transformed from the descriptive and cataloging field to a mechanistic one, whereby we have started to understand the molecular origins, processes, and mechanisms responsible for life of species. We dove deeper into molecular and atomic scale of these processes and learned how atomic constellations of biological molecules result in interactions between these molecules, resulting in higher-order physiological phenotypes. Not only we started to understand molecular mechanisms underlying physiological processes, but also learned how to rationally manipulate biological systems using genetic tools, molecular engineering, and drug discovery.

Molecular architecture and atomic structure have become central to biology. Recognizing the impact of structural biology, many nations contributed to Structural Genomics Initiative. The USA National Institutes of General Medical Sciences have sponsored one of the most successful programs, Protein Structure Initiative (PSI) [1], that resulted in significant expansion of coverage of protein fold space. This initiative has resulted in determination of nearly 7,000 protein structures including many unique folds. PSI has also promoted innovations in molecular structure determination methods and technologies [2].

Technological developments have been central to structural biology revolution. These developments can be stratified in to three categories: experimental, computational, and hybrid, in which sparse experimental, knowledge-based, and/or evolutionary information is used to determine protein structure. Next, we outline the strengths and limitations of these approaches, challenges, and strategies to tackle modern day problems of protein structure determination.

High-resolution experimental approaches

The solution of the first X-ray crystallographic structure of the protein myoglobin by Sir John Kendrew in 1958 [3] opened the flood gates of molecular structural revolution. Technological and scientific advances, that included the development of the nuclear magnetic resonance (NMR) techniques [4,5], negative staining and cryo-electron microscopy (cryo-EM) [6–8], computational and theoretical approaches [9–34], resulted in exponential explosion of the number of determined protein structures (Figure 1). In turn, newly discovered structures empowered scientists with unprecedented abilities to (i) establish structure-functional relationships in molecules, which resulted in a new field of molecular biology [35], (ii) discover new drugs based on small molecule-receptor structures, which resulted in the field of molecular pharmacology [36,37], (iii) design new proteins, which resulted in the field of computational protein design [38,39]. The structural revolution has had a profound impact on our society and we continue to benefit from the expanding field of structural biology.

Initial growth in determined protein structures has been due to advances in X-ray crystallography, however, the introduction of NMR for protein structure determination by Kurth Wüthrich [40–42] and others altered the landscape of molecular structure determination. Although one of the principal limitations of NMR has been the size of the molecule, a number of innovative techniques pushed the boundaries of molecular weight limit on studied proteins (Table 1). This technique offers a critical advantage with respect to crystallography, namely the ability to witness conformations of proteins, that float freely in, albeit crowded, solutions.

Technological innovations, as well as innovations in computer algorithms have led to breakthrough advances in cryo-EM technology and the rise of new era in structural biology. The number of protein structures solved using cryo-EM has been growing, including especially challenging transmembrane proteins and whole viruses [43]. According to estimates [44], we have not reached the theoretical limits of what is possible with cryo-EM, and the emerging field of cryo-electron tomography. Cryo-EM field promises to revolutionize molecular structure determination.

Computational modeling

Molecular modeling has been playing a critical role in structural determination. It is hard to imagine any structural biology experiment that would not rely on computational assistance in data processing and reconstruction. Molecular modeling has been used in molecular reconstruction of crystallographic data using molecular replacement, structural refinement, as well as building ensembles of molecular conformations using NMR data.

Ever increasing computational power enabled significant advances in molecular modeling. Initial success in structure prediction has come from phenomenological modeling, whereby known structural information was used to reconstruct previously unknown molecules. “Threading” [45] was one of the original dominant approaches for structure modeling that has been relying on a critical observation that if two proteins share sequence similarity, their structures are also similar (homologous) [46–50]. Thus, the first step in threading is

identification of homologous proteins (with known structures) to a given target protein. In the second step, all the sequence information is stripped from the identified homologous proteins and only backbone trace remains. In the third step, the sequence of the target protein is "threaded" through the backbone trace of the homologous proteins. Upon resolution of clashes and further structural optimization (energy minimization), the final energy of the threaded sequence, E, is evaluated and compared to that of the average of a randomly reshuffled sequences, $\bar{E}$ (to maintain amino acid composition). Based on this comparison one can build a distribution of Z-scores ($Z = \frac{E - \bar{E}}{\sigma_E}$, where σ_E is the standard deviation of the distribution). The probability, p, of finding a particular structure with a given Z-score can be obtained by integrating the tail of the normal distribution of Z-scores (due to the central limit theorem). This approach proved to be powerful [51–53], but limiting to closely related homologs. More distant homologous proteins featured backbone rearrangements. Since the threading score is susceptible to backbone variations, the threading procedure results in less accurate results for more distant homologs.

The next revolution in protein structure prediction has come from the fragment-based modeling, [54,55] whereby known protein structures are deconstructed into fragments of several amino acids. The library of fragments is then used to assemble the structures of an unknown target protein by matching corresponding sequence fragments of the target protein to that of a fragment from the library. This approach has proved to not only accurately predict protein structure, but enabled successful and robust protein design protocol, implemented in RosettaDesign [56]. Nonetheless, even this approach reached the limit of prediction accuracy as it is evidenced from the adaption of new approaches in the field of protein structure prediction.

The next breakthrough in computational structure determination was based on observation that if two amino acids in proteins form stabilizing interaction, then substitution of one would impact the other amino acid [57,58]. Based on this observation then, we can relate the strength of covariation of amino acids at various positions in the course of evolution with the chances that these residues interact in 3D space. Several methods have been developed to extract the potential interactions between positions along the chain using such direct coupling analysis [59–62], or DCA. As the result protein structure prediction reached new highs in the Critical Assessment of Structure Prediction (CASP) [63] competitions. The most recent breakthrough in computational structure prediction has come from the implementation of machine learning approaches [64–66]. Combination of the computational approaches has had a profound impact on structural biology.

Challenges

Despite these paramount advances in experimental and computational approaches, we are still facing some of the critical challenges in our understanding of molecular structure and dynamics. These challenges are typically associated with the complexity of characterization of intrinsic molecular dynamics and capturing conformational states of interest. Some examples of such obstacles include: (i) characterization of disordered proteins [67–72], (ii) identifying rare conformations [63–65], (iii) witnessing templating conformational conversions [73–76] (key steps in protein aggregation), (iv) determining structures of

molecular complexes, especially featuring interactions with disordered conformations [77–81], and (v) ligand-induced conformational disorder-order transitions [82–88]. All of these obstacles induce heterogeneity in samples, making structural determination extremely challenging.

The thread among above challenges is the characterization of disorder. Unlike structural determination of proteins that populate most of their life time in the near-native conformations, disordered proteins lack well-defined states and, instead, feature some persistent structural elements in otherwise diverse conformational ensemble. In this regard, the meaning of structural characterization is shifted from the concept of "one protein – one structure" to identifying persistent elements of conformational ensembles.

Problem formulation

How can we characterize these ensembles of disordered proteins? To understand how to characterize the ensembles, we look into the physical properties of heterogeneous proteins. Unlike random homopolymers, heteropolymers feature non-equal interactions between distinct regions of the polymers. According to Boltzmann distribution, the probability, $p(\Gamma)$, of observing a particular conformation, Γ, is $p(\Gamma) \sim exp\left(-\frac{E(\Gamma)}{k_B T}\right)$, where E is the energy of this conformation, k_B is the Boltzmann constant and T is the temperature. Thus, heteropolymeric segments that feature preferable interactions are more likely to appear in proximity within each other as such conformations feature lower energies than those that do not have strongly interacting segments in proximity. These more favorably interacting segments in proteins determine their conformational landscape. In structurally ordered proteins the attraction between interacting segments (ΔH) compete against conformational entropy (ΔS) to converge on a well-defined near-native structural ensemble, with stability $\Delta G = \Delta H - T\Delta S < 0$. In natively disordered proteins, there is little distinction between any sub-states and the unfolded states, i.e. $\Delta S \approx 0$. We can deconstruct contribution to $\Delta H = \Delta H_{sis} + \Delta H_{wis}$, where ΔH_{sis} is the enthalpic contribution of the strongly interacting segments, and ΔH_{wis} is that of the contribution of the weakly interacting segments. The probability to observe conformations with the strongly interacting segments present, considering that the weakly interacting segments have minimal energy contribution ($\Delta H_{wis} \approx 0$), would be $p_i(\Gamma) \sim exp\left(\frac{-\Delta H_{sis}(\Gamma) + T\Delta S}{k_B T}\right)$, which is much larger than that when these strongly interacting segments do not interact, $p_0(\Gamma) \sim exp\left(\frac{-\Delta H_{wis}(\Gamma) + T\Delta S}{k_B T}\right) : p_i(\Gamma) > p_0(\Gamma)$. Hence, in heteropolymers, one expects dynamic persistence of strongly interacting structural segments. Even in folded proteins, local unfolding around strongly interacting structural segments determine the dynamics of proteins and their aggregation morphologies [89].

Although strongly interacting segments shape protein conformational states, their characterization is not trivial as these conformational ensembles can be complex and context/condition-dependent. For example, there can be multiple competing conformations resulting in a given conformational ensemble. To better understand the problem, it is worth consider various evolutionary scenarios for appearance of natively disordered proteins.

There are several scenarios how proteins evolve to be disordered or feature some level of disorder (Figure 2). In one scenario, a stable protein loses its structure in the course of evolution. For example, partial loss of a function due to the stability loss of a given protein can be compensated to preserve fitness of the organism by having a compensation in a different protein with a similar function. Hence, the stability of the protein does not have to be maintained for organismal fitness. Yet unless a protein performs some function in cells, its expression becomes an energetic burden. Hence, proteins that lose their stability in the course of evolution either maintain its function, although at a lower level, or acquire new function(s). Not all mutations that result in disordered but functional protein are fixed. Thermodynamically stable proteins pack their hydrophobic cores on the inside and expose hydrophilic patches on the outside. When such proteins are unfolded, unfolded protein response (UPR) is initiated: these proteins are recognized by chaperones by binding to exposed hydrophobic patches. These chaperones either assist refolding of the misfolded protein or target it to proteasomal degradation. Hence, natively disordered proteins also evolve to evade UPR.

Another scenario for evolution of the natively disordered proteins may come from functional adaptation of a given protein. In the course of evolution, a protein may adapt a new function, and likely a new functional conformational ensemble. In this case, these two (original and acquired) functional ensembles compete with each other, making the protein *de facto* disordered, although it persistently populates these two functional states.

Although other scenarios for evolution are likely, the keys differentiating these scenarios are the number of dominant structural conformations and the kinetics of interconversion between them. Hence, to characterize the conformational states of the natively disordered proteins, one needs to generate these ensembles and, then, identify their chemical and biological properties, as well as the rates of conversions between states.

Molecular dynamics (MD) computer simulations allow generation of hypothetical ensembles of protein conformations. Upon sufficient sampling, MD simulations faithfully reproduce physical properties of natively disordered proteins [90]. However, sufficient sampling can be challenging with the traditional MD [91–100]. To circumvent this challenge, several approaches have been developed. One class of approaches is based on innovative strategies for sampling states, that include replica exchange (REX) [101,102] and accelerated MD (aMD) [103]. Another class of sampling enhancement approaches comes from parallelization, utilization of graphical processor units (GPUs) [104–107], and custom processor architectures, e.g. Anton [108]. Faster algorithms for simulations, such as Monte Carlo and discrete molecular dynamics (DMD) [109–113] offer increase in raw speed of simulations. These algorithms can often be combined with different sampling strategies, and hardware acceleration.

Even with software and/or hardware acceleration, the sampling maybe insufficient due to the size of the system. Furthermore, imperfections of the force fields may bias the conformational ensembles. A number of strategies have been developed to circumvent these challenges with the premise of restricting the search space by utilizing prior knowledge, such as evolutionary data, experiments, and literature. In all of these cases the restriction of

the search space impacts the entropy and reduces the free energy barrier that separates low energy conformational state(s) from the random coil states. Next, we will describe several such strategies.

Fitting the experimental collective observables

Various experiments offer different types of observables. In case when experiments report on collective variables, S^i_{exp}, for a given observable i, a method to restrict computational sampling of molecular conformations is to impose constraints on the difference between computational, S^i_{comp}, and experimental values of the collective observables by minimizing their $\chi^2\left(\overrightarrow{S}_{comp}, \overrightarrow{S}_{exp}\right)$ values:

$$\begin{cases} \chi^2\left(\overrightarrow{S}_{comp}, \overrightarrow{S}_{exp}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(S^i_{comp} - S^i_{exp}\right)^2 \\ \chi^2_{min} = min_{\overrightarrow{S}_{comp}}\chi^2\left(\overrightarrow{S}_{comp}, \overrightarrow{S}_{exp}\right) \end{cases} \quad (1)$$

The solution of these equations guarantees an optimal ensemble that satisfies experimental data and the level of agreement between thus computationally-constructed and experimental conformational ensembles is characterized by the value of χ^2_{min}.

This approach has been extensively used in protein folding community to build protein folding-unfolding transition state ensembles by using as the collective variables ϕ-values, which is an approximate measure of a particular residue, i, contribution to the transitional state ensemble [114–121]: $S^i = \phi_i$. Mapping transition state ensembles by fitting of the experimental ϕ-values allowed the determination of the structural properties of some of the kinetically most evasive states.

Sample and select

If sampling of conformational states is not an obstacle, an alternative strategy, named “sample and select”, is based on the generation of the naïve (without experimental biases) conformational ensembles in MD simulations, and selection of conformations from the obtained ensembles that satisfy experimental data using Equation (1). This approach has been developed by Chen et al. [122], for utilizing NMR residual dipolar coupling (RDC) data as the bias to obtain conformational ensembles consistent with the RDCs. It has also been utilized by other groups for interpreting NMR data of RNA and DNA molecules [123–125].

Pairwise constraints-based modeling

Various experiments, such as amino-acid cross linking, FRET, and NMR, and evolutionary inference studies [57–62] often report on pairwise proximity of residues. Pairwise proximity data can be readily incorporated into the physical force field, H_{phys},

$$H = H_{phys} + \lambda H_{constr}, \quad (2)$$

where

$$H_{constr} = \sum_{(i,j)_{constr}} U_{ij}(r), \quad (3)$$

$(i, j)_{constr}$ are pairs of constrained residues, and $\lambda < 1$ is a weight coefficient for the constraints' potential. The form of potentials, $U_{ij}(r)$, can be chosen as spring (quadratic) potential, if it is used in traditional MD engine, or square wells, if used in DMD (Figure 3). The middle of the well, r_{ij}, corresponds to the average distance expected between residues i and j, and the width, σ_{ij}, corresponds to the expected variation of distances between these residues.

An important experimental limitation that needs to be considered is that some of the pairwise proximity data set may conflict within itself. The solution to this limitation is described in the *Ensemble deconstruction* and *Conflicting and "dirty" constraints* sections below.

The utilization of the pairwise proximity information offers a rapid construction of protein conformations that are consistent with this information [126]. Satisfying each constraint reduces entropy of protein by $\Delta S \sim ln|i - j|$ [127], thus driving the conformational ensemble towards the states that satisfy constraints and minimize the free energy of the system.

A number of groups develop methodologies for incorporating cross-linking and co-evolutionary constraints [128–130]. Borchers and Dokholyan laboratories developed cross-linking driven DMD engine that streamlines utilization of chemical cross-linking information in DMD simulation [72,131] (Figure 4). This approach has been successfully applied to a number of proteins [131,132], including intrinsically disordered proteins (IDP) [133–135] α-synuclein [72] and tau [71] proteins.

Shape constraint-based modeling

Shape of the protein can be determined using several approaches, such as small angle X-ray scattering (SAXS), atomic force microscopy (AFM), and negative staining EM. Shape can be also used as constraints for building a structure. Many packages utilizing SAXS data build molecular shapes by filling in spheres into the volume and optimizing the shape of the volume until significant match between experimental and computational intensities of scattered X-ray [136–140].

Alternative approaches may include: (i) fitting the shape density profiles to computationally-derived ones by using χ^2 minimization, similar to Equation (1); (ii) determining shape fingerprints, such as Zernike functions, and matching experimentally derived ones to those coming from known structures [141,142]; (iii) using simulations with biasing potential towards a particular shape. While the last approach has not been implemented for *ab initio* structure prediction, Dokholyan laboratory uses this approach routinely to bias simulations using experimental constraints [79].

Surface exposure-based modeling

Measuring exposure of various amino acids to solvent can be readily achieved in biochemical and biophysical experiments. Next, we describe several techniques used to perform protein surface mapping.

In a limited proteolysis approach a protein is subject to proteases for a period of time sufficient for a single cut to occur. The conditions are controlled by the buffer and the reaction is rapidly quenched at the time point where we expect a single proteolytic event. This approach requires calibration of protease reaction kinetics through a set of time course experiments. Single cuts are most likely to occur on the surface of a thermodynamically-stable proteins, while cleavable sites are buried in the protein cores. The cleavage sites are readily detectable using mass spectrometry (MS).

In these experiments, thermodynamic stability of a protein determines how noisy is the data. For a two-state protein the ratio of folded versus unfolded state is determined by the equilibrium constant $K_F = {}^{f}/_{u}$, where f and u are fraction of folded and unfolded states correspondingly. If $K_F \gg 1$, then proteolytic cleavage occurs predominantly in the folded state. In contrast, when $K_F \sim 1$, cleavage of the unfolded conformations contributes to the signal, thereby mixing folding and unfolding states. Even for marginally stable proteins, limited proteolysis can be a viable approach to determine structures of protein native folded states, although in this case, quantitative MS should be used to determine contribution of each states and deconvolute them using *Ensemble deconstruction* technique described below.

An important benefit of the proteolysis is that some of the proteases are specific to a particular amino acid sequence epitope. If these proteases do not cleave a protein in a particular specific to these proteases' sites, this negative information can also be used as a constraint in simulations.

An approach to utilize surface exposure data in simulations was proposed by Proctor et al. [143], who developed a modified Gō potential [144,145] to perform DMD simulations. Gō force field biases the protein towards the native state by assigning attractive or repulsive potential to pairs of residues that are proximal or distal in the native state [67,146,147]. In this approach, experimental data was used as a biasing potential to a native Gō potential [110,144,145], i.e.

$$\mathrm{H} = -\epsilon_0 \sum_{i<j} \Delta_{ij}\delta_{ij} + \lambda \sum_{l \in cuts} \sum_{k<l} \mathrm{E}_{cut}(j,k), \tag{4}$$

where the first term is the Gō potential and the second term is the biasing term that is attractive for all potential cut sites that are not cut in experiments, i.e. they are protected, and is repulsive for the potential cut sites that were cut in experiments [143]. Using this potential, Proctor et al, was able to propose a model of a transient trimeric intermediate of superoxide dismutase 1 (SOD1) that appears on its aggregation pathway and has relevance to amyotrophic lateral sclerosis (ALS), discussed below.

Hydrogen-deuterium exchange (HDX) experiments utilize the propensity for amide hydrogens to be replaced by deuteriums when protein is immersed into the heavy water. The

exchange rates can be measured by NMR (quantitatively) in the EX2 limit (in which conformational changes occur at faster than the exchange between hydrogen and deuterium) [148] and MS. In the EX2 regime, satisfied for most stable proteins, the protection factor of a given residue, P_f, is a ration of the intrinsic k_{rc} and an experimentally measured k_{ex} exchange rates, $P_f = {k_{rc}}/{k_{ex}}$. Protection factors are related to residues (k) contributions to the free energy differences between open and closed states:

$$\Delta G_{HDX} = \sum_k \Delta G_{HDX}(k) = -RT \sum_k \ln P_f(k). \quad (5)$$

There are two approaches to reconstruct protein conformational ensembles using protection factors: one was developed by Vendruscolo et al. [149], and one developed in Dokholyan laboratory [148]. Vendruscolo et al. proposed to minimize the difference between the protection factors determined from the simulated conformational ensembles and experimental ones, in essence, similar to Equation (1). Dokholyan laboratory used protection factors to design a force field that would guide rapid DMD simulations towards ensembles consistent with the observed protection factors. The latter approach is agnostic about physical force field since we generate ensembles based on the force field derived directly from the experiments.

Chemical modifications mapping is another approach to utilize residue reactivity to reagents present in solutions. The modifications of residues exposed to the solvent that reacted with the reagents, can be detected using MS or chemical/physical sensors, such as fluorescent reporter binding at the modification site[150]. The approach to model molecular structure based on chemical modification is conceptually similar to the approaches used for HDX or limited proteolysis-based structure modeling.

Ensemble deconstruction

All described methods result in conformational ensembles, Γ, that are consistent with various experimental data. These ensembles contain information about (meta)stable states, some of them are rare. To uncover these states, clustering of protein conformations is performed based on the pairwise root-mean square distance (RMSD) between all conformations. The clustering procedure partitions the conformational ensemble into states, Γ_k, whose population is the weight of the cluster, w_k:

$$\Gamma \sim \cup_{k=1}^{N_{cl}} \Gamma_k, \quad (6)$$

where N_{cl} is the number of obtained clusters. In essence, these clusters represent conformational landscape of a protein [151,152] and the logarithm of weights is related to the free energy of these states:

$$F_k = -k_B T \log w_k, \quad (7)$$

Where k_B is the Boltzman constant and T is the temperature.

Depending on the algorithm, the clustering may depend on one or more free parameter that, in most cases, is/are chosen *ad hoc*. This threshold or cutoff parameter defines the number of clusters N_{cl} that one expects to obtain, which requires intuition about the properties of system under consideration. One algorithm, where there is no arbitrary cutoff parameter is based on the physical assumption of criticality, i.e. the partitioning of the protein conformational space is a critical phenomenon. In such case, the number of clusters in clustering critically depends on the order parameter, which is the cutoff. The critical point of this transition is at the midpoint of quazi-first order or second-order phase transition. In the first order phase transition critical points separates two discontinuous phases, and the transition is discontinuous. However, due to the finite size of the system, the boundary conditions do not allow discontinuity and the dependence of N_{cl} as a function of the cutoff is typically a sharp sigmoidal curve (Figure 6). In the second-order phase transition, N_{cl} depends on the cutoff also as a sigmoidal curve, although this curve is typically less sharp at the transition point than that in the quasi-first order transition. The advantage of this clustering approach is that it does not require subjective intuition about the number of states in the system. Criticality-based clustering was implemented by Dokholyan et al. [153] to uncover the intrinsic evolutionary relationship between structures of unrelated proteins.

Conflicting and “dirty” constraints

Experimental measurements come with errors stemming from (i) the instrumental inaccuracies, (ii) incorrect interpretation/assignment of measured values (wrong assignment), and, (iii) when dealing with molecular conformations, measurements of variables, that due to insufficient instrumental resolution, effectively represent an average between distinct states (insufficient resolution). Next, we will address methods for mitigating these errors in computational modeling using experimental constraints using protocols established over years in Dokholyan’s laboratory using DMD simulations [110–112,146,147,154–158], although these protocols can be implemented in other simulation packages.

(i) Mitigating instrumental inaccuracies. A hallmark feature of the DMD simulations is the force field that is represented by square-well potentials. In essence, the force field itself can be thought of a set of pairwise constraints between different atom types, $U^{Medusa}(r)$, (in all atom force field, e.g. Medusa) [159]. Integration of experimental pairwise constraints is then just an addition of the square well potentials, U_i^{cons}, to the main force field with the width, δ, of the square well corresponding to the instrumental variability of the signal (Figure 3).

$$U(r) = U^{medusa}(r) + w\sum\nolimits_i U_i^{cons}(r), \tag{8}$$

where w is the relative weight of the constraints and

$$U_i^{cons}(r) = \begin{cases} -\varepsilon_i, d_i - \delta_i/2 < r < d_i + \delta_i/2 \\ 0, otherwise \end{cases}. \tag{9}$$

The depth of the square well, ε, is set to contribute comparatively with other terms of the force field. For simplicity, all the well depth parameters ε_i can be set to a single value ε,

unless there is evidence for other experimental biases. The distances d_i between atoms or amino acids are derived from the experimental constraints.

How to choose the weight, w, in DMD simulation? The answer to this question is in the physical properties of a protein. The energy function Eq. (8) defines the folding temperature, hence, the weight is chosen in such a way that the melting temperature of the protein agrees with the experimental value. Hence, the weight should be adjusted in such a way that (a) the constraints terms $\sum_i U_i^{cons}(r)$ do not overpower the physical terms $U^{Medusa}(r)$, i.e. both terms are of the same order of magnitude, and (b) the folding transition temperature matches experimental values. Alternatively, weight w can be chosen by minimizing the square difference of observable parameters measured in simulations and experiments, and identifying the value of w_{min}, where this difference is minimal.

(ii) Wrong assignments may appear due to human or algorithmic errors in assigning the experimental data to a particular atom/residue(s).

(iii) Insufficient resolution. Lack of experimental resolution may result in overlap/averaging between two or more distinct states and measured value corresponds to a non-physical conformation. For example, if the time scale resolution of experiments is longer than the life time of a protein in either of two states, the measured variable will represent an average between these two states, which may not have any physical meaning (Figure 7).

Both errors (ii) and (iii) are automatically mitigated by the sufficient sampling and the sufficient amount of properly assigned constraints. Clustering typically separates conformational states. Constraints stemming from both wrong assignments and the insufficient resolution would form substates distinguishable from other states because non-physical constraints effectively frustrations in the free energy landscape of proteins [160–162]. These frustrations can be mitigated by extensive sampling that overcomes artificial free energy barriers introduced by these constraints. Thus, identified conformational states can either be physical or artificial. Even if the incorrect experimental constraints dominate the system, a small number of correct constraints may be sufficient to “pull down” physical states [163].

How to distinguish physical states? Any computational study results in a model, and, as such, this model requires validation. Hence, a critical step in any modeling studies is a *forward validation* step, whereby based on the developed model, one predicts previously-unknown properties of the system. With the advent of protein design, this approach presents a viable and potent tool to perform forward validation of the model protein. For example, in order to determine transient oligomeric states of SOD1, Proctor et al. [143] implemented surface exposure-based modeling by performing limited proteolysis of the isolated oligomers SOD1 and identifying the proteolytic sites using MS. Identified cut sites shape a map of solvent accessibility, which can be used to build model structures of oligomers. For forward validation, Proctor et al. designed mutations that stabilize and destabilize the oligomers, which indeed had corresponding effects in both biochemical and cellular assays.

Additional consideration to keep in mind when applying constraints is that molecular kinetics may play role in how imposed constraints are satisfied, i.e. even if all constraints are representing a single state, the order of the constraints' satisfaction may have a detrimental impact on whether the molecule can reach the target state. Hence, the constraints should be "soft" enough to allow them to form and break easily, if the kinetics of folding prevents satisfaction of all constraints (Figure 3CD).

Ensemble completeness

How can we make sure that the computational sampling is sufficient? The real answer is never. Ultra-rare state may never be sampled in simulations. Perhaps the most pragmatic question is: when do we know that we have achieved sufficient enough conformational sampling to address particular biological questions? One approach to address this question, is to compare the properties of measured in simulation observables ω derived from trajectories spanning time t and $2t$. Indistinguishability of these observables, $\omega(t) \approx \omega(2t)$, indicates a plausible convergence of the simulation and an adequate sampling. If these observables vary significantly, $\omega(t) \neq \omega(2t)$, then further simulation time doubling is performed until convergence.

Integrated modeling

Complex problems often require complex solutions. Building structural models of multifaceted molecules and molecular complexes may require multiple diverse orthogonal and nonorthogonal experimental interrogations of these molecules. In this complex system, integrating developed algorithms to incorporate different types of constraints offers a direct approach for building structural models. Perhaps one of the most striking examples of integrated model building was developed by Sali laboratory [164], where they were able to build one of the largest complexes in living cells, 50 MDa nuclear pore complex (NPC), which comprises of 456 proteins [165]. Alber et al. integrated a number of different experiments, ranging from crystallography, NMR, cryo-EM, affinity purification and computational modeling to reconstruct NPC [166]. Sali laboratory offers *imp* package for integrated modeling [137].

Multiscale modeling

Computational or experimental approaches offer only a limited view of the biological processes in living cells, which can span a broad range of time and length scales. For example, studies of electron transport require understanding quantum processes, conformational dynamics, and protein-protein interactions. Over the past two decades many multiscale modeling approaches have been developed that allow spanning sufficiently broad range of the life of a single biological molecule. Many laboratories are moving towards developing multiscale modeling for whole cell simulations. The pioneering work from Schulten's laboratory [167] demonstrated the feasibility of such approaches. Conceptually, the multiscale modeling workflow consists of approaches that span overlapping but distinct time and length scale regions. These approaches must agree in the overlap regions, and these regions tend to be particularly challenging [158] to model due to inherently different methods used in these approaches. Nobel Prize in Chemistry 2013 (The Nobel Prize in Chemistry 2013. NobelPrize.org) was awarded to Michael Levitt, Ariel Warshell, and Martin

Karplus for the development of multiscale approaches in computational chemistry. Since their original pioneering works, a number of computational approaches have been developed to span scales relevant to life of biological molecules [26,168]. Constraintsbased modeling can be incorporated into multiscale modeling workflow at any scale, barring in mind effects those constraints may have on the boundary regions between different scales.

Resolution of computational models

The resolution of structural model is a critical parameter in experimental structure determination, as it allows one to understand the level of "trust" of the determined atomic positions. Likewise, the resolution of computational model is as critical. One approach to evaluate the resolution of a structural model that is used in Dokholyan laboratory is to relate uncertainty in atomic positions from the fluctuations of the molecular structure within a low free energy basin, i.e. resolution of the model is approximated by the root mean square fluctuations over the ensemble of conformations. We perform simulated annealing simulations [169,170], whereby the system gradually is brought to a free energy basin with satisfied experimental constraints. At this point, the root mean square fluctuations of the conformational ensemble determines allowed space for our system, thereby determining the resolution of the structural model. This approached was proposed by Chen et al. [171], and Dokholyan laboratory has used it for internal evaluation of structural models. For example, in attempt to understand the molecular etiology of cystic fibrosis, we build a structural model of the cystic fibrosis transmembrane conductance regulator (CFTR), which revealed that the most common mutation ΔF508 occurs at the interface between the surface of the nucleotide-binding domain 1 (NBD1) and a cytoplasmic loop (CL4) in the C-terminal membrane-spanning domain (MSD2) [172]. We further validated the model by performing cross-linking experiments, which tested the proximity of residues predicted by our structural model. The experiments indeed validated our model. To determine the accuracy of CFTR in the NBD1-CL4 interface, we performed simulations with the constrains imposed by the linkers and, using a protocol described above, determined that our resolution is approximately 3 Å (unpublished). This approach can be broadly utilized for estimating resolutions of structural models based on experimental constraints.

Future challenges

Approaches to model protein structure are becoming increasingly complex and are being integrated into multi-step workflows. Each of the approaches has limitations and the integration often requires explicitly taking those limitations in the consideration. The advantage of these modeling workflows is that they offer one stop solution for building molecular structures. The drawback is that these modeled structures are taking without further validation. Constructions of CFTR [172,173] ryanodine receptor (RyR) [174–177], and the dynein [178,179] models have been particularly challenging because direct modeling workflows failed and we had to resort to manual molecular structure reconstruction. Understanding limitations of each of the step in the modeling workflow and forward validation of computational models in experiments are critical to validity of the molecular models.

The advent of machine learning techniques offers new horizons in our ability not only to build structural models, but also to select models most likely to be resembling “real” proteins. The advantage of the machine learning approaches is in establishing connections in data, otherwise “invisible” to rational thought processes. The disadvantage in such approaches is our inability to gauge whether the constructed structural models are appropriate and truthful representations of the targets. Thus, forward experimental validations of structural models built using machine learning approaches are essential to the validity of molecular models.

Acknowledgements

We thank Drs. Venkat Chirasani and Jian Wang for their help with the manuscript preparation and Bradley M. Winters for his help with illustrations. We acknowledge support from the National Institutes for Health (5R01GM123247, 2R01 GM114015, and 1R35 GM134864 to NVD) and the Passan Foundation. The project described was also supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR002014. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH.

References

1. Chandonia J-M, Brenner SE. The Impact of Structural Genomics: Expectations and Outcomes. Science (80-). 2006 1;311(5759):347 LP – 351.
2. Schwede T, Sali A, Honig B, Levitt M, Berman HM, Jones D, et al. Outcome of a workshop on applications of protein models in biomedical research. Structure. 2009 2;17(2):151–9. [PubMed: 19217386]
3. KENDREW JC, BODO G, DINTZIS HM, PARRISH RG, WYCKOFF H, PHILLIPS DC. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. Nature. 1958;181(4610):662–6. [PubMed: 13517261]
4. Rabi II, Zacharias JR, Millman S, Kusch P. A New Method of Measuring Nuclear Magnetic Moment. Phys Rev. 1938 2;53(4):318.
5. Filler A. The History, Development and Impact of Computed Imaging in Neurological Diagnosis and Neurosurgery: CT, MRI, and DTI. Internet J Neurosurg. 2010 5;7:1–85.
6. Henderson R, Unwin PN. Three-dimensional model of purple membrane obtained by electron microscopy. Nature. 1975 9;257(5521):28–32. [PubMed: 1161000]
7. Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J Mol Biol. 1990 6;213(4):899–929. [PubMed: 2359127]
8. Cheng Y, Grigorieff N, Penczek PA, Walz T. A Primer to Single-Particle Cryo-Electron Microscopy. Cell. 2015;161(3):438–49. [PubMed: 25910204]
9. Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973 7;181(4096):223–30. [PubMed: 4124164]
10. Wetlaufer DB. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci U S A. 1973 3;70(3):697–701. [PubMed: 4351801]
11. Dill KA. Dominant forces in protein folding. Biochemistry. 1990 8;29(31):7133–55. [PubMed: 2207096]
12. Klimov Thirumalai. Criterion that determines the foldability of proteins. Phys Rev Lett. 1996 5;76(21):4070–3. [PubMed: 10061184]
13. Guo Z, Brooks CL 3rd. Thermodynamics of protein folding: a statistical mechanical study of a small all-beta protein. Biopolymers. 1997 12;42(7):745–57. [PubMed: 10904547]
14. Lazaridis T, Karplus M. “New view” of protein folding reconciled with the old through multiple unfolding simulations. Science. 1997 12;278(5345):1928–31. [PubMed: 9395391]
15. Koppensteiner WA, Sippl MJ. Knowledge-based potentials--back to the roots. Biochemistry (Mosc). 1998 3;63(3):247–252. [PubMed: 9526121]

16. Baldwin RL. Structure and mechanism in protein science. A guide to enzyme catalysis and protein folding, by A. Fersht. 1999. New York: Freeman. 631 pp. $67.95 (hardcover). Protein Sci. 2000 1;9(1):207.
17. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M, Šali A, et al. Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci. 2000 7;25(7):331–9. [PubMed: 10871884]
18. Nelson Onuchic J, Nymeyer H, García AE, Chahine J, Socci NDBT-A in PC. The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios In: Protein folding mechanisms. Academic Press; 2000 p. 87–152.
19. Koehl P, Levitt M. Theory and simulation. Can theory challenge experiment? Vol. 9, Current opinion in structural biology. England; 1999 p. 155–6. [PubMed: 10465610]
20. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels, and protein function. Protein Sci. 1999 6;8(6):1181–90. [PubMed: 10386868]
21. Go N. Theoretical studies of protein folding. Annu Rev Biophys Bioeng. 1983;12(1):183–210. [PubMed: 6347038]
22. Scala A, Dokholyan NV, Buldyrev SV, Stanley HE. Thermodynamically important contacts in folding of model proteins. Phys Rev E Stat Nonlin Soft Matter Phys. 2001 3;63(3 Pt 1):32901.
23. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. Proc Natl Acad Sci. 2002 6;99(13):8637 LP – 8641. [PubMed: 12084924]
24. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev E Stat Nonlin Soft Matter Phys. 2002 6;65(6 Pt 1):61910.
25. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory. Part I: Basic concepts. Q Rev Biophys. 2002 5;35(2):111–67. [PubMed: 12197302]
26. Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol. 2004;14(1):70–5. [PubMed: 15102452]
27. Matysiak S, Clementi C. Mapping folding energy landscapes with theory and experiment. Arch Biochem Biophys. 2008 1;469(1):29–33. [PubMed: 17910943]
28. Dill KA. Theory for the folding and stability of globular proteins. Biochemistry. 1985 3;24(6):1501–9. [PubMed: 3986190]
29. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J Phys Chem. 1989 9;93(19):6902–15.
30. Privalou PL. Thermodynamic Problems of Protein Structure. Annu Rev Biophys Biophys Chem. 1989 6;18(1):47–69. [PubMed: 2660833]
31. Daggett V, Levitt M. A model of the molten globule state from molecular dynamics simulations. Proc Natl Acad Sci U S A. 1992 6;89(11):5142–6. [PubMed: 1594623]
32. Karplus M, Šali A. Theoretical studies of protein folding and unfolding. Curr Opin Struct Biol. 1995;5(1):58–73. [PubMed: 7773748]
33. Shakhnovich EI. Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol. 1997;7(1):29–40. [PubMed: 9032061]
34. Shakhnovich E, Abkevich V, Ptitsyn O. Conserved residues and the mechanism of protein folding. Nature. 1996 1;379(6560):96–8. [PubMed: 8538750]
35. Consortium GO. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330–8. [PubMed: 30395331]
36. Jorgensen WL. Computer-aided discovery of anti-HIV agents. Bioorg Med Chem. 2016;24(20):4768–78. [PubMed: 27485603]
37. Marrone TJ, Briggs James M and, McCammon JA. Structure-based drug design: computational advances. Annu Rev Pharmacol Toxicol. 1997;37(1):71–90. [PubMed: 9131247]
38. Kuhlman B, Bradley P. Advances in protein structure prediction and design. Nat Rev Mol Cell Biol. 2019;20(11):681–97. [PubMed: 31417196]
39. Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. Nature. 2016;537(7620):320–7. [PubMed: 27629638]
40. Wuthrich K, Wagner G. NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor. FEBS Lett. 1975 2;50(2):265–8. [PubMed: 234403]

41. Williamson MP, Havel TF, Wuthrich K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. J Mol Biol. 1985 3;182(2):295–315. [PubMed: 3839023]
42. Qian YQ, Billeter M, Otting G, Muller M, Gehring WJ, Wuthrich K. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell. 1989 11;59(3):573–80. [PubMed: 2572329]
43. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol. 2008;18(5):581–6. [PubMed: 18674618]
44. Glaeser RM. How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits? Annu Rev Biophys. 2019 5;48(1):45–61. [PubMed: 30786229]
45. Rost B, Schneider R, Sander C. Protein fold recognition by prediction-based threading. J Mol Biol. 1997 7;270(3):471–80. [PubMed: 9237912]
46. Chothia C, Lesk A. The relation between the divergence of sequence and structure in proteins. EMBO J [Internet]. 1986;5(4):823–6. Available from: http://pku.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnZ3PS8MwFMcfczDcRXQ6nD8gp902lyZpGtgGMjZEFDzoxctIm4QNtS37cdC_3iRth716KfSlKeGl9L0XPnwfAFLMJa6Sa6KN4jJh0kVBHTLGOdZeuvj9OXp6CRZz9tiAChlzRGWNSxym65VnK0snbnxztaFxRaIzqQnGYRiFrL_Lss9J_uXn6Z9-8aBXHZKxOzF [PubMed: 3709526]
47. Levitt M. Accurate modeling of protein conformation by automatic segment matching. J Mol Biol. 1992 7;226(2):507–33. [PubMed: 1640463]
48. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct. 2000;29:291–325. [PubMed: 10940251]
49. Baker D, Sali A. Protein structure prediction and structural genomics. Science. 2001 10;294(5540):93–6. [PubMed: 11588250]
50. Zhang Y, Skolnick J. The protein structure prediction problem could be solved using the current PDB library. Proc Natl Acad Sci U S A. 2005/01/14. 2005 1;102(4):1029–34. [PubMed: 15653774]
51. Russell RB, Saqi MA, Bates PA, Sayle RA, Sternberg MJ. Recognition of analogous and homologous protein folds--assessment of prediction success and associated alignment accuracy using empirical substitution matrices. Protein Eng. 1998;11(1):1–9. [PubMed: 9579654]
52. Koretke KK, Russell RB, Copley RR, Lupas AN. Fold recognition using sequence and secondary structure information. Proteins Struct Funct Bioinforma. 1999;37(S3):141–8.
53. Zhou H, Zhou Y. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. Proteins Struct Funct Bioinforma. 2004;55(4):1005–13.
54. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol [Internet]. 1997;268(1):209–25. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=9149153&query_hl=1 [PubMed: 9149153]
55. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. Methods Enzymol. 2004;383:66–93. [PubMed: 15063647]
56. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science. 2003 11;302(5649):1364–8. [PubMed: 14631033]
57. Gobel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. Proteins. 1994 4;18(4):309–17. [PubMed: 8208723]
58. Shindyalov IN, Kolchanov NA, Sander C. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? Protein Eng. 1994 3;7(3):349–58. [PubMed: 8177884]

59. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein–protein interaction by message passing. Proc Natl Acad Sci. 2009 1;106(1):67 LP – 72. [PubMed: 19116270]
60. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. Protein 3D structure computed from evolutionary sequence variation. PLoS One. 2011;6(12):e28766. [PubMed: 22163331]
61. Schug A, Weigt M, Onuchic JN, Hwa T, Szurmant H. High-Resolution Protein Complexes from Integrating Genomic Information with Molecular Simulation. Proc Natl Acad Sci U S A. 2009;106(52):22124–9. [PubMed: 20018738]
62. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci U S A. 2011 12;108(49):E1293–301. [PubMed: 22106262]
63. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)-Round XII. Proteins. 2018 3;86 Suppl 1:7–15. [PubMed: 29082672]
64. Gao M, Zhou H, Skolnick J. DESTINI: A deep-learning approach to contact-driven protein structure prediction. Sci Rep. 2019;9(1):3514. [PubMed: 30837676]
65. AlQuraishi M. End-to-End Differentiable Learning of Protein Structure. bioRxiv. 2018 1;265231.
66. Hou J, Wu T, Cao R, Cheng J. Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. Proteins Struct Funct Bioinforma. 2019 4;0(0).
67. Dokholyan NV. Studies of folding and misfolding using simplified models. Curr Opin Struct Biol. 2006;16(1):79–85. [PubMed: 16413773]
68. Krzeminski M, Marsh JA, Neale C, Choy W-Y, Forman-Kay JD. Characterization of disordered proteins with ENSEMBLE. Bioinformatics. 2013 2;29(3):398–9. [PubMed: 23233655]
69. Delaforge E, Cordeiro TN, Bernadó P, Sibille N. Conformational Characterization of Intrinsically Disordered Proteins and Its Biological Significance BT - Modern Magnetic Resonance. In: Webb GA, editor. Cham: Springer International Publishing; 2017 p. 1–20.
70. Gibbs EB, Showalter SA. Quantitative Biophysical Characterization of Intrinsically Disordered Proteins. Biochemistry. 2015 2;54(6):1314–26. [PubMed: 25631161]
71. Popov KI, Makepeace KAT, Petrotchenko EV, Dokholyan NV, Borchers CH. Insight into the Structure of the “Unstructured” Tau Protein. Structure. 2019;
72. Brodie NI, Popov KI, Petrotchenko EV, Dokholyan NV, Borchers CH. Conformational ensemble of native α-synuclein in solution as determined by short-distance crosslinking constraint-guided discrete molecular dynamics simulations. PLOS Comput Biol. 2019 3;15(3):e1006859. [PubMed: 30917118]
73. Ding F, LaRocque JJ, Dokholyan NV. Direct observation of protein folding, aggregation, and a prion-like conformational conversion. J Biol Chem. 2005;280(48):40235–40. [PubMed: 16204250]
74. Makarava N, Baskakov IV. Genesis of tramsmissible protein states via deformed templating. Prion. 2012 7;6(3):252–5. [PubMed: 22561163]
75. Zhou Z, Xiao G. Conformational conversion of prion protein in prion diseases. Acta Biochim Biophys Sin (Shanghai). 2013 6;45(6):465–76. [PubMed: 23580591]
76. Baskakov IV. Switching in amyloid structure within individual fibrils: implication for strain adaptation, species barrier and strain classification. FEBS Lett. 2009/05/29. 2009 8;583(16):2618–22. [PubMed: 19482025]
77. Wang Q, Chen M, Schafer NP, Bueno C, Song SS, Hudmon A, et al. Assemblies of calcium/calmodulin-dependent kinase II with actin and their dynamic regulation by calmodulin in dendritic spines. Proc Natl Acad Sci. 2019;116(38):18937–42. [PubMed: 31455737]
78. Zhu C, Dukhovlinova E, Council O, Ping L, Faison EM, Prabhu SS, et al. Rationally designed carbohydrate-occluded epitopes elicit HIV-1 Env-specific antibodies. Nat Commun [Internet]. 2019;10(1):948 Available from: http://www.ncbi.nlm.nih.gov/pubmed/30814513 [PubMed: 30814513]
79. Dagliyan O, Proctor EAA, D’Auria KMM, Ding F, Dokholyan NVV. Structural and Dynamic Determinants of Protein-Peptide Recognition. Structure [Internet]. 2011 12 7 [cited 2017 May

9];19(12):1837–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22153506 [PubMed: 22153506]

80. Furukawa Y, Teraguchi S, Ikegami T, Dagliyan O, Jin L, Hall D, et al. Intrinsic Disorder Mediates Cooperative Signal Transduction in STIM1. J Mol Biol [Internet]. 2014 5;426(10):2082–97. Available from: http://www.sciencedirect.com/science/article/pii/S0022283614001302 [PubMed: 24650897]
81. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PW. Simulating disorder–order transitions in molecular recognition of unstructured proteins: Where folding meets binding. Proc Natl Acad Sci. 2003 4;100(9):5148 LP – 5153. [PubMed: 12697905]
82. Dokholyan NV. Controlling Allosteric Networks in Proteins. Chem Rev [Internet]. 2016 6 8 [cited 2017 May 9];116(11):6463–87. Available from: 10.1021/acs.chemrev.5b00544 [PubMed: 26894745]
83. Dagliyan O, Shirvanyants D, Karginov AV, Ding F, Fee L, Chandrasekaran SN, et al. Rational design of a ligand-controlled protein conformational switch. Proc Natl Acad Sci [Internet]. 2013;110(17):6800–4. Available from: http://www.pnas.org/content/110/17/6800.short [PubMed: 23569285]
84. Dagliyan O, Tarnawski M, Chu P-H, Shirvanyants D, Schlichting I, Dokholyan NV, et al. Engineering extrinsic disorder to control protein activity in living cells. Science (80-) [Internet]. 2016 12 16 [cited 2017 May 9];354(6318):1441 LP – 1444. Available from: http://science.sciencemag.org/content/354/6318/1441.abstract
85. Dagliyan O, Krokhotin A, Ozkan-Dagliyan I, Deiters A, Der CJ, Hahn KM, et al. Computational design of chemogenetic and optogenetic split proteins. Nat Commun [Internet]. 2018;9(1):4042 Available from: 10.1038/s41467-018-06531-4 [PubMed: 30279442]
86. Dagliyan O, Dokholyan NV, Hahn KM. Engineering proteins for allosteric control by light or ligands. Nat Protoc [Internet]. 2019;14(6):1863–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31076662 [PubMed: 31076662]
87. Pang X, Zhou H-X. Disorder-to-Order Transition of an Active-Site Loop Mediates the Allosteric Activation of Sortase A. Biophys J. 2015 10;109(8):1706–15. [PubMed: 26488662]
88. Zea DJ, Monzon AM, Gonzalez C, Fornasari MS, Tosatto SCE, Parisi G. Disorder transitions and conformational diversity cooperatively modulate biological function in proteins. Protein Sci. 2016 6;25(6):1138–46. [PubMed: 27038125]
89. Ding F, Furukawa Y, Nukina N, Dokholyan NV. Local unfolding of Cu, Zn superoxide dismutase monomer determines the morphology of fibrillar aggregates. J Mol Biol. 2012;421(4–5):548–60. [PubMed: 22210350]
90. Ding F, Jha RK, Dokholyan NV. Scaling behavior and structure of denatured proteins. Structure. 2005 7;13(7):1047–54. [PubMed: 16004876]
91. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem. 1983 6;4(2):187–217.
92. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009 7;30(10):1545–614. [PubMed: 19444816]
93. MacKerell AD Jr., Brooks B, Brooks CL III, Nilsson L, Roux B, Won Y, et al. CHARMM: The Energy Function and Its Parameterization. Encyclopedia of Computational Chemistry. 1998 (Major Reference Works).
94. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. Wiley Interdiscip Rev Comput Mol Sci. 2013;3(2):198–210.
95. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KMJ, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005 12;26(16):1668–88. [PubMed: 16200636]
96. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701–18. [PubMed: 16211538]
97. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput [Internet]. 2008

3;4(3):435–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26620784 [PubMed: 26620784]

98. Kutzner C, Van Der Spoel D, Fechner M, Lindahl E, Schmitt UW, De Groot BL, et al. Speeding up parallel GROMACS on high-latency networks. J Comput Chem. 2007 9;28(12):2075–84. [PubMed: 17405124]
99. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005 12;26(16):1781–802. [PubMed: 16222654]
100. Plimpton S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. J Comput Phys. 1995;117(1):1–19.
101. Swendsen RH, Wang J-S. Replica Monte Carlo Simulation of Spin-Glasses. Phys Rev Lett. 1986 11;57(21):2607–9. [PubMed: 10033814]
102. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999;314(1–2):141–51.
103. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004 6;120(24):11919–29. [PubMed: 15268227]
104. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. J Comput Chem. 2007 12;28(16):2618–40. [PubMed: 17894371]
105. Anderson JA, Lorenz CD, Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units. J Comput Phys. 2008;227(10):5342–59.
106. Yasuda K. Accelerating Density Functional Calculations with Graphics Processing Unit. J Chem Theory Comput. 2008 8;4(8):1230–6. [PubMed: 26631699]
107. Harger M, Li D, Wang Z, Dalby K, Lagardere L, Piquemal J-P, et al. Tinker-OpenMM: Absolute and relative alchemical free energies using AMOEBA on GPUs. J Comput Chem. 2017 9;38(23):2047–55. [PubMed: 28600826]
108. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, et al. Anton, a Special-purpose Machine for Molecular Dynamics Simulation. Commun ACM. 2008 7;51(7):91–7.
109. Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford university press; 2017.
110. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des [Internet]. 1998 11 [cited 2017 Aug 30];3(6):577–87. Available from: http://pku.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwrV1LT8MwDI5gF7jw2JgYDyknBIfSNm3aRJomjT2EEEhIwIVL1CatVm3rpo0duPLLcZp2jNMkxNFWmrROZX92bAchnIKK44p4EedUxsSTgJqd1Jepm4YAb3Uw__2JPT6T4YA-lHUwZUZlqfaNOi8UdcmxS0Ha8yyzX1xPdx0JGWfaxhFd6xvAorrLY_9urYhDh5o [PubMed: 9889167]
111. Shirvanyants D, Ding F, Tsao D, Ramachandran S, Dokholyan NV. Discrete Molecular Dynamics: An Efficient And Versatile Simulation Method For Fine Protein Characterization. J Phys Chem B [Internet]. 2012 7 26 [cited 2017 May 9];116(29):8375–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22280505 [PubMed: 22280505]
112. Proctor EA, Ding F, Dokholyan NV. Discrete molecular dynamics. Wiley Interdiscip Rev Comput Mol Sci. 2011 1;1(1):80–92.
113. Rapaport DC. The Art of Molecular Dynamics Simulation. 2nd ed Cambridge: Cambridge University Press; 2004.
114. Matouschek A, Kellis JT, Serrano L, Bycroft M, Fersht AR. Transient folding intermediates characterized by protein engineering. Nature. 1990;346(6283):440–5. [PubMed: 2377205]
115. Matouschek A, Kellis JT, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. Nature. 1989;340(6229):122–6. [PubMed: 2739734]
116. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. J Mol Biol. 2000 5;298(5):937–53. [PubMed: 10801360]

117. Nymeyer H, Socci ND, Onuchic JN. Landscape approaches for determining the ensemble of folding transition states: Success and failure hinge on the degree of frustration. Proc Natl Acad Sci. 2000 1;97(2):634 LP – 639. [PubMed: 10639131]
118. Ozkan SB, Bahar I, Dill KA. Transition states and the meaning of Phi-values in protein folding kinetics. Nat Struct Biol. 2001 9;8(9):765–9. [PubMed: 11524678]
119. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. Nature. 2001 2;409(6820):641–5. [PubMed: 11214326]
120. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Identifying the protein folding nucleus using molecular dynamics. J Mol Biol. 2000 3;296(5):1183–8. [PubMed: 10698625]
121. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Direct molecular dynamics observation of protein folding transition state ensemble. Biophys J. 2002 12;83(6):3525–32. [PubMed: 12496119]
122. Chen Y, Campbell SL, Dokholyan NV. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. Biophys J [Internet]. 2007 10 1;93(7):2300–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17557784 [PubMed: 17557784]
123. Torchia DA. NMR studies of dynamic biomolecular conformational ensembles. Prog Nucl Magn Reson Spectrosc. 2015;84–85:14–32.
124. Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, et al. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat Chem Biol. 2011;7(8):553. [PubMed: 21706033]
125. Ying J, Grishaev A, Bax A. Carbon-13 chemical shift anisotropy in DNA bases from field dependence of solution NMR relaxation rates. Magn Reson Chem. 2006 3;44(3):302–10. [PubMed: 16477676]
126. V S Ali A, Blundell TL, Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993 12;234(3):779–815. [PubMed: 8254673]
127. Doi M. Introduction to polymer physics. Oxford university press; 1996.
128. dos Santos RN, Ferrari AJR, de Jesus HCR, Gozzo FC, Morcos F, Martínez L. Enhancing protein fold determination by exploring the complementary information of chemical cross-linking and coevolutionary signals. Bioinformatics [Internet]. 2018;34(13):2201–8. Available from: 10.1093/bioinformatics/bty074 [PubMed: 29447388]
129. Ferrari AJR, Gozzo FC, Martínez L. Statistical force-field for structural modeling using chemical cross-linking/mass spectrometry distance constraints. Bioinformatics [Internet]. 2019;35(17):3005–12. Available from: 10.1093/bioinformatics/btz013 [PubMed: 30629125]
130. Belsom A, Schneider M, Fischer L, Brock O, Rappsilber J. Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. Mol Cell Proteomics [Internet]. 2015; Available from: https://www.mcponline.org/content/early/2015/09/18/mcp.M115.048504
131. Brodie NI, Popov KI, Petrotchenko EV, Dokholyan NV, Borchers CH. Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations. Sci Adv [Internet]. 2017 7 1;3(7):e1700479 Available from: http://advances.sciencemag.org/content/3/7/e1700479.abstract [PubMed: 28695211]
132. Petrotchenko EV, Xiao K, Cable J, Chen Y, Dokholyan NV, Borchers CH. BiPS, a photocleavable, isotopically coded, fluorescent cross-linker for structural proteomics. Mol Cell Proteomics. 2009 2;8(2):273–86. [PubMed: 18838738]
133. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem. 2014;83:553–84. [PubMed: 24606139]
134. Uversky VN. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. Vol. 7, Frontiers in Physics. 2019 p. 10.
135. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2014 12;16:18.
136. Panjkovich A, Svergun DI. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. Bioinformatics. 2018 6;34(11):1944–6. [PubMed: 29300836]

137. Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol. 2012 1;10(1):e1001244. [PubMed: 22272186]

138. Doniach S, Lipfert J. Use of small angle X-ray scattering (SAXS) to characterize conformational states of functional RNAs. Methods Enzymol. 2009;469:237–51. [PubMed: 20946792]

139. Liu H, Hexemer A, Zwart PH. The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. J Appl Crystallogr. 2012;45(3):587–93.

140. Baul U, Chakraborty D, Mugnai ML, Straub JE, Thirumalai D. Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins. J Phys Chem B [Internet]. 2019 4 25;123(16):3462–74. Available from: 10.1021/acs.jpcb.9b02575 [PubMed: 30913885]

141. Venkatraman V, Yang YD, Sael L, Kihara D. Protein-protein docking using region-based 3D Zernike descriptors. BMC Bioinformatics. 2009 12;10:407. [PubMed: 20003235]

142. Yin S, Dokholyan NV. Fingerprint-based structure retrieval using electron density. Proteins. 2011 3;79(3):1002–9. [PubMed: 21287628]

143. Proctor EA, Fee L, Tao Y, Redler RL, Fay JM, Zhang Y, et al. Nonnative SOD1 trimer is toxic to motor neurons in a model of amyotrophic lateral sclerosis. Proc Natl Acad Sci [Internet]. 2015 1 19 [cited 2017 May 9];113(3):614–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26719414 [PubMed: 26719414]

144. Ueda Y, Taketomi H, Go N, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific inter-unit interactions. Int J Pept Protein Res. 1975;7(6):445–59. [PubMed: 1201909]

145. GŌ N, Abe H. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. Biopolymers. 1981 5;20(5):991–1011. [PubMed: 7225531]

146. Ding F, Dokholyan NV. Simple but predictive protein models. Trends Biotechnol. 2005;23(9):450–5. [PubMed: 16038997]

147. Proctor EA, Dokholyan NV. Applications of Discrete Molecular Dynamics in biology and medicine. Curr Opin Struct Biol [Internet]. 2016 4 [cited 2017 May 9];37:9–13. Available from: 10.1016/j.sbi.2015.11.001 [PubMed: 26638022]

148. Dixon RDS, Chen Y, Ding F, Khare SD, Prutzman KC, Schaller MD, et al. New insights into FAK signaling and localization based on detection of a FAT domain folding intermediate. Structure [Internet]. 2004 12;12(12):2161–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15576030 [PubMed: 15576030]

149. Vendruscolo M, Paci E, Dobson CM, Karplus M. Rare Fluctuations of Native Proteins Sampled by Equilibrium Hydrogen Exchange. J Am Chem Soc. 2003 12;125(51):15686–7. [PubMed: 14677926]

150. Aprahamian ML, Chea EE, Jones LM, Lindert S. Rosetta protein structure prediction from hydroxyl radical protein footprinting mass spectrometry data. Anal Chem. 2018;90(12):7721–9. [PubMed: 29874044]

151. Panchenko AR, Luthey-Schulten Z, Wolynes PG. Foldons, protein structural modules, and exons. Proc Natl Acad Sci. 1996;93(5):2008–13. [PubMed: 8700876]

152. Panchenko AR, Luthey-Schulten Z, Cole R, Wolynes PG. The foldon universe: a survey of structural similarity and self-recognition of independently folding units. J Mol Biol. 1997;272(1):95–105. [PubMed: 9299340]

153. Dokholyan NV, Shakhnovich B, Shakhnovich EI. Expanding protein universe and its origin from the biological Big Bang. Proc Natl Acad Sci. 2002 10;99(22):14132 LP – 14136. [PubMed: 12384571]

154. Ding F, Buldyrev SV, Dokholyan NV. Folding Trp-Cage to NMR Resolution Native Structure Using a Coarse-Grained Protein Model. Biophys J. 2005;88(1):147–55. [PubMed: 15533926]

155. Chen Y, Ding F, Nie H, Serohijos AW, Sharma S, Wilcox KC, et al. Protein folding: then and now. Arch Biochem Biophys. 2008;469(1):4–19. [PubMed: 17585870]

156. Ding F, Tsao D, Nie H, Dokholyan NV. Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. Structure. 2008;16(7):1010–8. [PubMed: 18611374]

157. Ding F. Discrete Molecular Dynamics Simulation of Biomolecules. Comput Model Biol Syst From Mol to Pathways, Biol Med Physics, Biomed Eng. 2011 10;

158. Sparta M, Shirvanyants D, Ding F, Dokholyan NV, Alexandrova AN. Hybrid Dynamics Simulation Engine for Metalloproteins. Biophys J. 2012;103(4):767–76. [PubMed: 22947938]

159. Ding F, Dokholyan NV. Emergence of Protein Fold Families through Rational Design. PLOS Comput Biol [Internet]. 2006 7 7;2(7):e85 Available from: 10.1371/journal.pcbi.0020085 [PubMed: 16839198]

160. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG. Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery. Proc Natl Acad Sci U S A. 2005 2;102(7):2373 LP – 2378. [PubMed: 15701699]

161. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins. 1995 3;21(3):167–95. [PubMed: 7784423]

162. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem. 1997;48:545–600. [PubMed: 9348663]

163. Benfeard Williams II, Zhao B, Tandon A, Ding F, Weeks KM, Zhang Q, et al. Structure modeling of RNA using sparse NMR constraints. Nucleic Acids Res [Internet]. 2017 12 15;45(22):12638–47. Available from: 10.1093/nar/gkx1058 [PubMed: 29165648]

164. Alber F, Forster F, Korkin D, Topf M, Sali A. Integrating diverse data for structure determination of macromolecular assemblies. Annu Rev Biochem. 2008;77:443–77. [PubMed: 18318657]

165. Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, Raveh B, et al. Integrative structure and functional anatomy of a nuclear pore complex. Nature. 2018 3;555:475. [PubMed: 29539637]

166. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, et al. The molecular architecture of the nuclear pore complex. Nature. 2007 11;450(7170):695–701. [PubMed: 18046406]

167. Perilla JR, Schulten K. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. Nat Commun. 2017;8(1):15959. [PubMed: 28722007]

168. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. J Phys Chem B. 2012 7;116(29):8494–503. [PubMed: 22545654]

169. Sharma S, Ding F, Nie H, Watson D, Unnithan A, Lopp J, et al. iFold: a platform for interactive folding simulations of proteins. Bioinformatics. 2006 11;22(21):2693–4. [PubMed: 16940324]

170. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science (80-). 1983 5;220(4598):671–80.

171. Chen Y, Ding F, Dokholyan NV. Fidelity of the protein structure reconstruction from inter-residue proximity constraints. J Phys Chem B [Internet]. 2007 6 28;111(25):7432–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17542631 [PubMed: 17542631]

172. Serohijos AWR, Hegedus T, Aleksandrov AA, He L, Cui L, Dokholyan NV, et al. Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. Proc Natl Acad Sci U S A [Internet]. 2008 3 4;105(9):3256–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18305154 [PubMed: 18305154]

173. Hegedus T, Serohijos AWR, Dokholyan NV, He L, Riordan JR. Computational studies reveal phosphorylation-dependent changes in the unstructured R domain of CFTR. J Mol Biol. 2008 5;378(5):1052–63. [PubMed: 18423665]

174. Ramachandran S, Serohijos AWR, Xu L, Meissner G, Dokholyan NV. Ryanodine receptor pore structure and function. Biophys J. 2009;96(3):107a.

175. Xu L, Mowrey DD, Chirasani VR, Wang Y, Pasek DA, Dokholyan NV, et al. G4941K substitution in the pore-lining S6 helix of the skeletal muscle ryanodine receptor increases RyR1 sensitivity to cytosolic and luminal Ca 2+. J Biol Chem. 2018 2;293(6):2015–28. [PubMed: 29255089]

176. Xu L, Chirasani VR, Carter JS, Pasek DA, Dokholyan NV, Yamaguchi N, et al. Ca2+-mediated activation of the skeletal-muscle ryanodine receptor ion channel. J Biol Chem. 2018 12;293(50):19501–9. [PubMed: 30341173]

177. Ramachandran S, Chakraborty A, Xu L, Mei Y, Samsó M, Dokholyan NV, et al. Structural Determinants of Skeletal Muscle Ryanodine Receptor Gating. J Biol Chem. 2013 3;288(9):6154–65. [PubMed: 23319589]

178. Serohijos AWR, Chen Y, Ding F, Elston TC, Dokholyan NV. A structural model reveals energy transduction in dynein. Proc Natl Acad Sci U S A. 2006/11/22. 2006 12;103(49):18540–5. [PubMed: 17121997]

179. Serohijos AWR, Tsygankov D, Liu S, Elston TC, Dokholyan NV. Multiscale approaches for studying energy transduction in dynein. Phys Chem Chem Phys. 2009 6;11(24):4840–50. [PubMed: 19506759]

Significance

Experimentally-driven computational structure modeling and determination is a rapidly evolving alternative to traditional approaches for molecular structure determination. These new hybrid experimental-computational approaches are proving to be a powerful microscope to glance into the structural features of intrinsically or partially disordered proteins, dynamics of molecules and complexes. In this review, we describe various approaches in the field of experimentally-driven computational structure modeling.

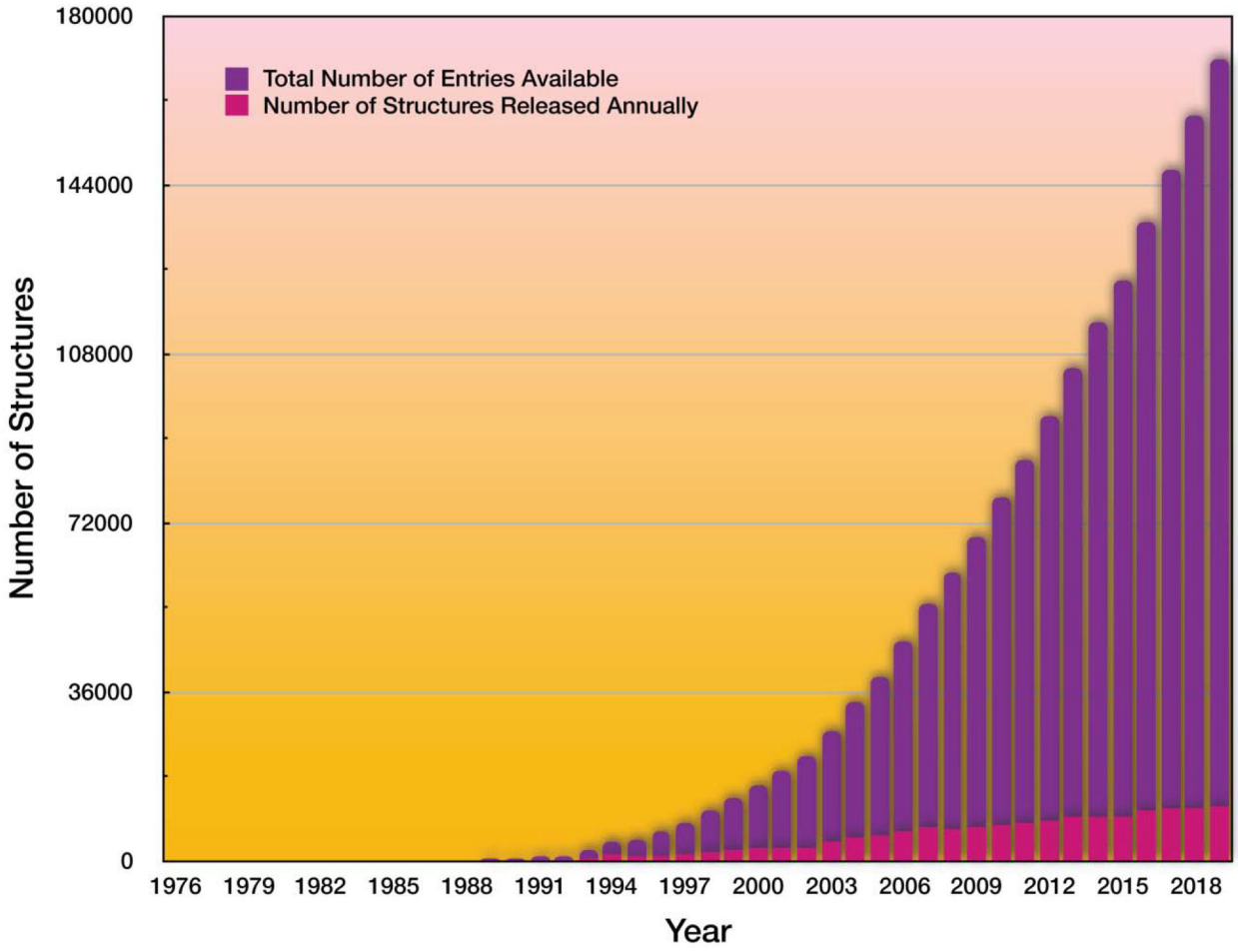


Figure 1.
Exponential knowledge growth of structures of biological molecules (mostly proteins) over the past 40 years.

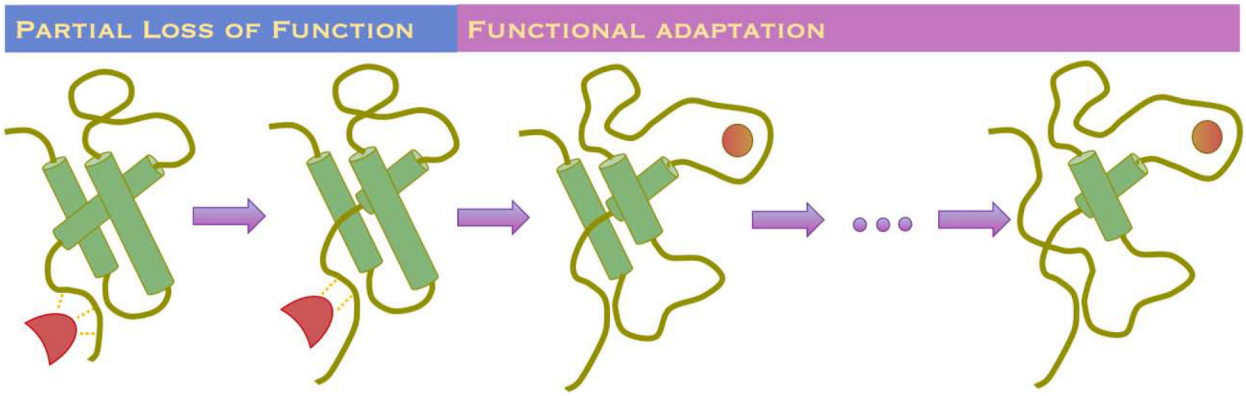


Figure 2. Two evolutionary scenarios resulting in functional yet intrinsically disordered protein.
In one scenario, a partial loss of function can be compensated by the increased functionality of orthologous genes or reduced need of a given functionality in evolved environments. In another scenario, a given scaffold can be adapted to carry a distinct function that require less ordering of the protein.

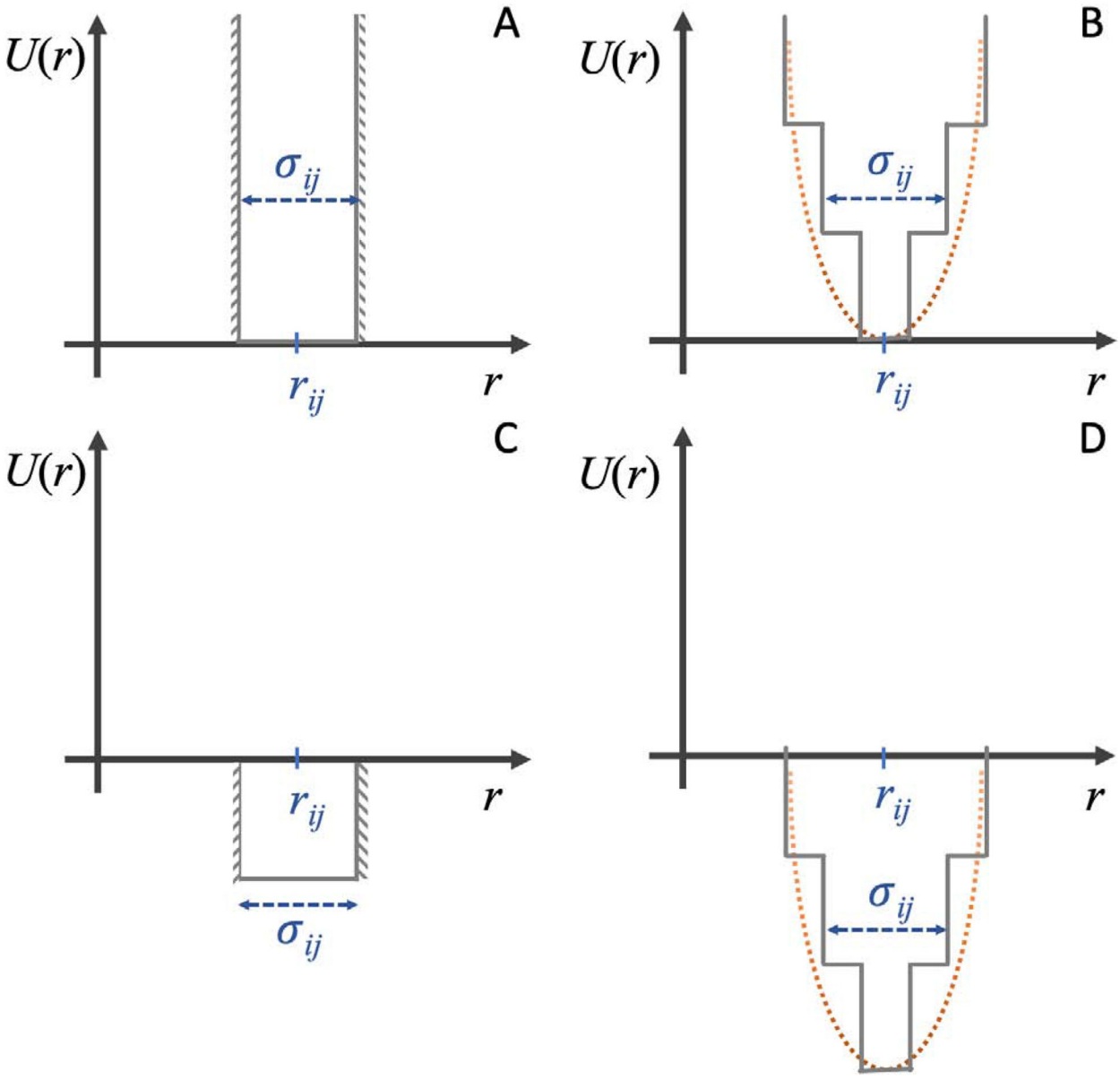


Figure 3. Implementations of constrains in the force field.

(A) Restrictive constraints enforce residues within a given range. (B) Hookean spring potential is a “softer” analog of potential (A). This potential can be approximated by square-well potential function to enable its implementation in DMD. (C) Biasing constraints make it more favorable for residues to be within a certain range. This potential allows constraints to not be satisfied, which is an important property when constraints, representing distinct states, are in conflict with each other. (D) Hybrid constraints potential between that of (B) and (C).

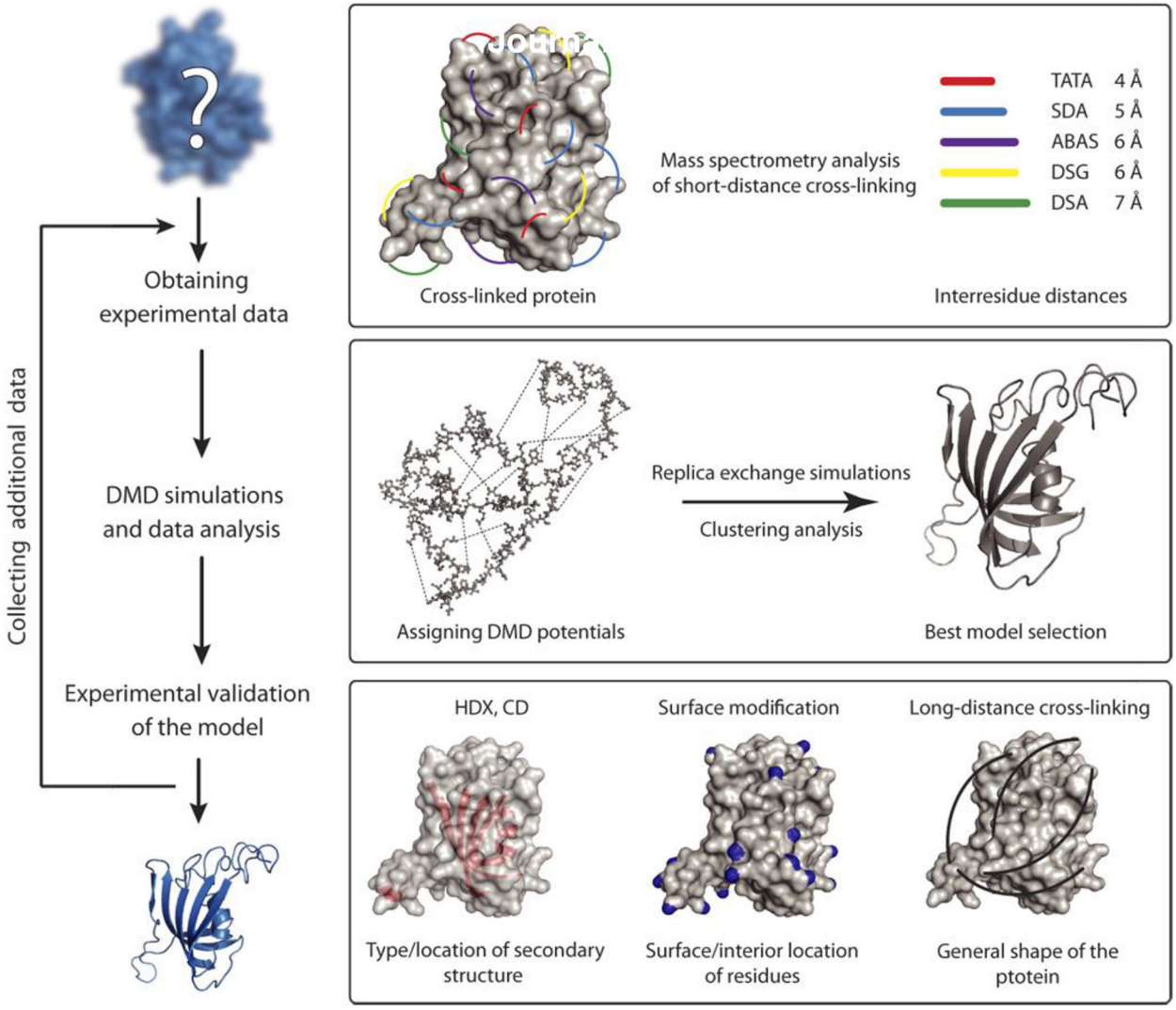


Figure 4. The workflow for CL-DMD protein structure prediction [117].
Copyright (2017) Science.

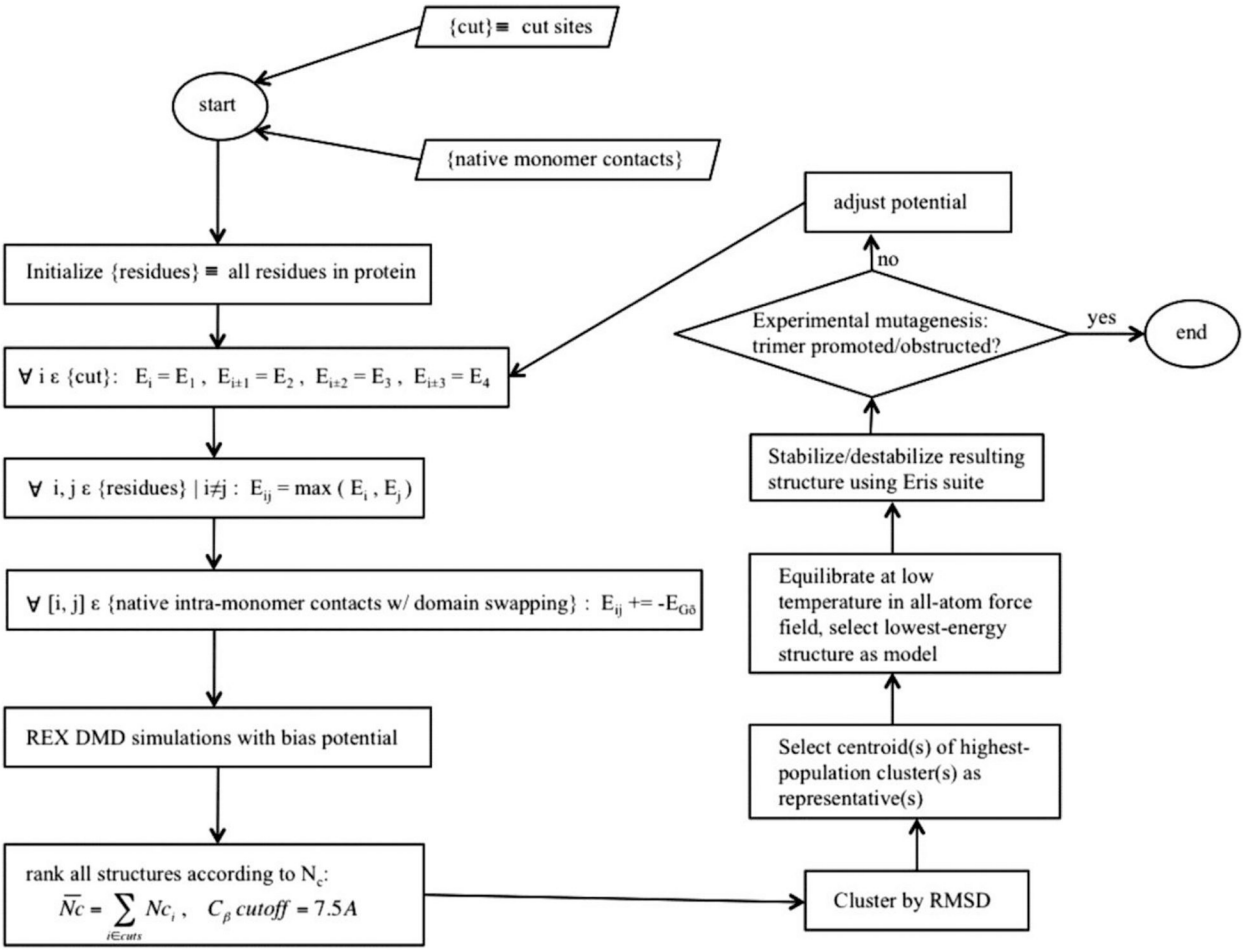


Figure 5. The workflow for structural modeling of protein complexes using information derived from proteolytic cleavage of these complexes [143].

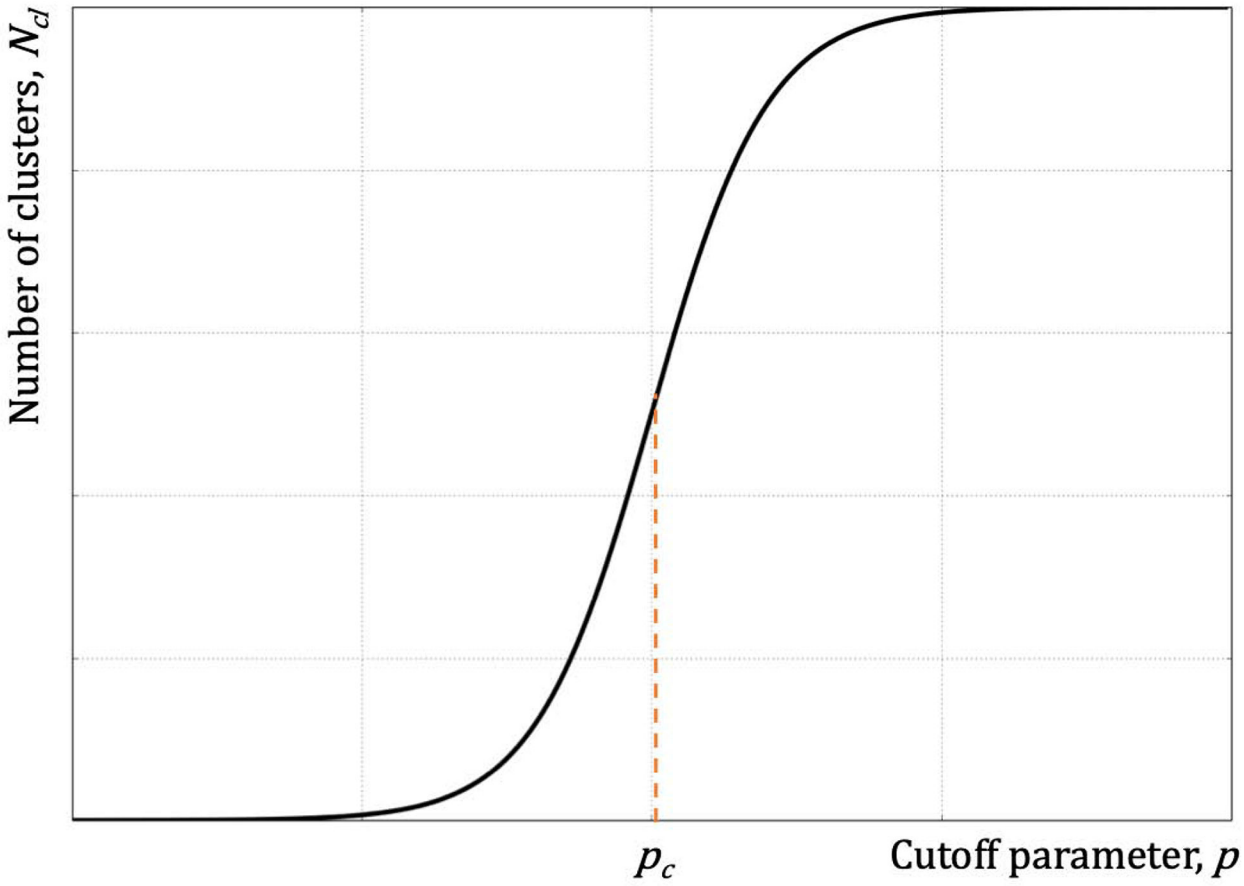


Figure 6. The dependence of the size of the largest cluster on the cutoff parameter, *p*, is a sigmoidal curve.

The midpoint of the transition, p_c, denotes the critical regime with the appearance of clusters of all sizes with one maintaining the majority of the elements.

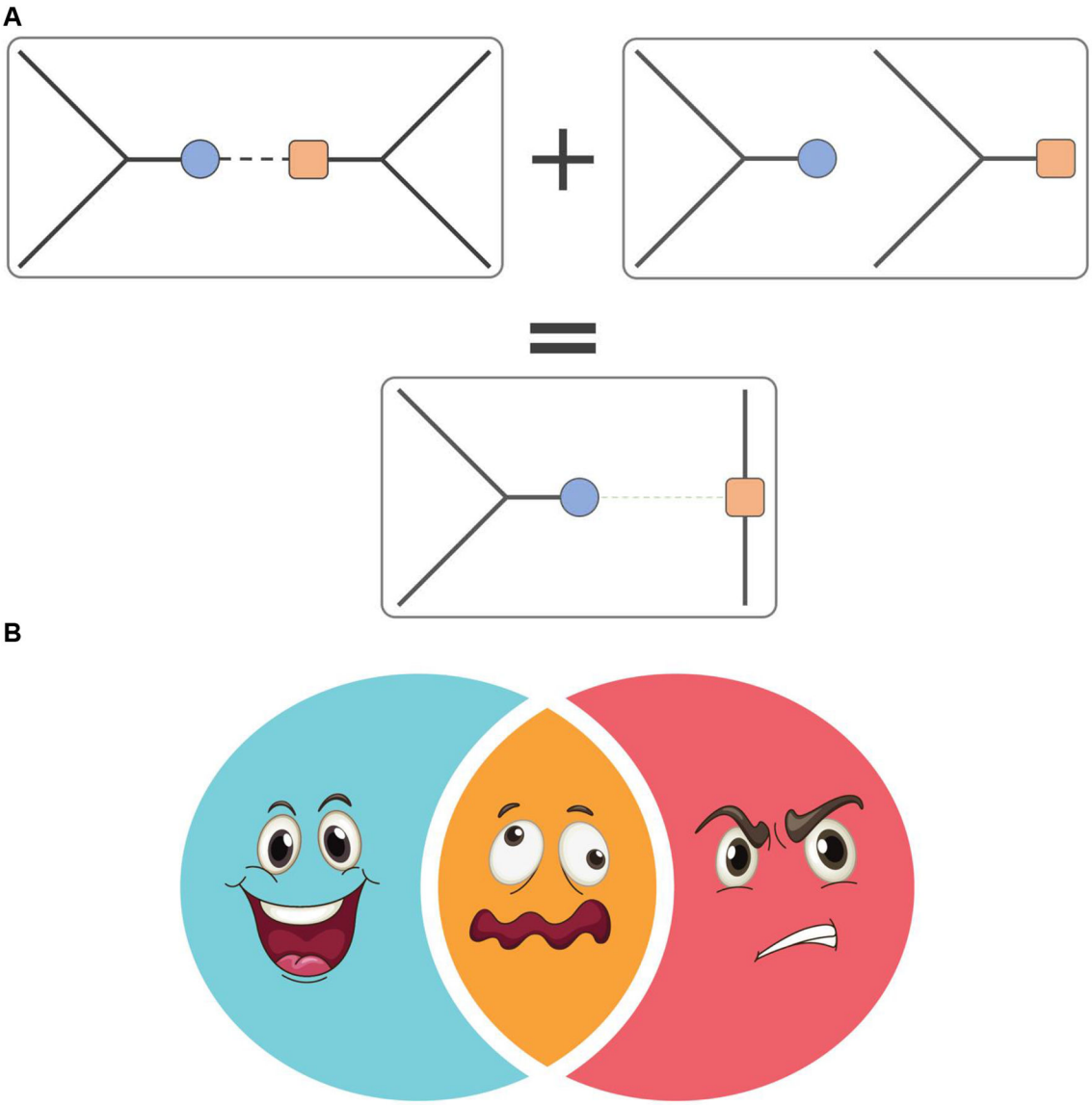


Figure 7. Superposition of two distinct physical states may have no real physical manifestation.
(A) In this example, the peptide bond of the average state is non-physical (bottom) upon averaging structures of two physical states of the ensemble (top). (B) Superposition of happy and angry faces does not result in a neutral one, but rather in not a face.

Table 1.

Advantages and limiations of experimental structure determination techniques.

Technique	**Advantages**	**Limitations**
X-ray crystallography	Large proteins and complexes High resolution structures	Requires high protein concentration Possible crystallographic artifacts: packing, non-native conformatins *Ad hoc* detemination of crystalizing conditions for the target protein Limited to stable proteins Molecular motion is challenging to obtain
NMR	Multiple complementary techniques for structure determination: two (and higher)-dimensional NMR, residual dipolar coupling, hydrogen-deuterium exchange, paramagnetic relaxation enhancements Ability to quantify dynamics and witness large-scale conformational dynamics of proteins	Requires high protein concentration Atominc assignment may present challenges
CryoEM	Large structures Ability to witness many conformational states Fairly rapid workflow	Small proteins still present a challenge